2013

Validation of Copy Number Variants Associated with Schizophrenia Risk in an Irish Population and Implications to Clinical Practice

Rachel L. Elves
Virginia Commonwealth University

Validation of Copy Number Variants Associated with Schizophrenia Risk in an Irish Population and Implications to Clinical Practice

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

by

Rachel Leigh Elves
Master of Sciences (M.Sc.)
University of British Columbia, Canada, 2008

Director: Rita Shiang, Ph.D.
Associate Professor, Department of Human and Molecular Genetics

Virginia Commonwealth University
Richmond, Virginia
August 2013

Acknowledgements

I would like to thank my family and friends for their wonderful support, including my parents, my husband Rob, and Amy Adkins, Laura Hack, and Jia Yan. A big thank you to my extended Richmond family, Ja'Neil Jackson, and Buddy and Nancy Corbett, for looking out for me and for giving me great memories to take with me on my journey. Special thanks to my advisors Rita Shiang, Sarah Elsea, and John Quillin for all of their help down this long road, and to the Riley lab for their help with my project and for the use of their lab space and equipment.

Table of Contents

List of Tables

List of Figures

Abstract

VALIDATION OF COPY NUMBER VARIANTS ASSOCIATED WITH SCHIZOPHRENIA RISK IN AN IRISH POPULATION AND IMPLICATIONS TO CLINICAL PRACTICE

By Rachel Leigh Elves, M.Sc.

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

Virginia Commonwealth University, 2013

Director: Rita Shiang, Ph.D.

Associate Professor, Department of Human and Molecular Genetics

Schizophrenia is a complex disorder affecting 1% of the population and is highly heritable, but the majority of contributing genetic factors has remained elusive. Current risk estimates for clinical practice are primarily determined by family history and associated empirical risk. Copy number variants (CNVs) may hold the key to explaining the missing

heritability in schizophrenia research; schizophrenia risk estimates as high as 30% have been found for the most-studied CNV associated with schizophrenia, 22q11.

Currently, there are methods to identify CNVs though previously collected data from SNP microarrays that would facilitate these types of studies. To determine if algorithms that call CNVs from microarray data are robust four genomic regions with putative CNVs called by the Wellcome Trust Consortium using Birdseye in Birdsuite with Affymetrix 6.0 array raw SNP intensities, primarily affecting genes *CHD1L*, *COX5B*, *PAK7*, *ZFYVE20*, were validated using Taqman real-time qPCR assays in 29 samples by research groups at VCU and Dublin. CNVs called from the algorithm were 100% validated at VCU though there were false negatives from the algorithm that were validated. Two samples at loci with putative duplications were not called by the Dublin group, which may be because of differing sensitivities of the Taqman assays to be able to detect a 50% difference in copy number between duplications and diploid controls, or because of another technical or analytical difference between the two sites.

Deletion frequency of one common CNV found in the gene *ERBB4*, was assessed by qPCR in both Irish singleton (ICCSS) and Irish family (IHDSF) samples and compared with Irish control (Trinity Biobank) and North American control populations. The *ERBB4* deletion frequency was not significantly different when comparing the Irish controls to the Irish singleton or the Irish family samples though the family samples were different when compared against the North American control population, which suggests population stratification, rather than a true association between *ERBB4* and increased schizophrenia risk.

Current clinical practice has been improved by knowledge and evaluation of CNV-related disorders that include risk for psychosis and additional phenotypes. Genotyping of individuals with known psychosis has led to improved patient care for non-psychosis-related phenotypes

associated with CNVs. Individuals with suspected genomic disorders that are found to have CNVs can be counseled on potential psychosis risk and potential risk to their offspring. Recurrent CNVs may hold promise in future clinical practice in order to individualize risk estimates in the general patient population, and increase the number of individuals able to receive anticipatory treatment to minimize disease severity.

Chapter 1: Background and Literature Review

Introduction

Schizophrenia is a complex neurodevelopmental disorder that strikes in the prime of life between the ages of 18 and 35, and affects as many as 24 million individuals worldwide[1,2]. Up to 50% of individuals are not receiving adequate treatment for their symptoms[3] that can include: hallucinations, delusions, disorganized speech and behavior, and catatonia[1]. Perhaps more concerning is negative symptoms, like flattened mood, loss of motivation, and reduced or absent speech. Negative symptoms are resistant to treatment and can lead to withdrawal from society[3]. This effect is compounded by the stigma faced by individuals resulting in isolation and potential worsening of symptoms[2,3].

Neurological and Behavioral Characteristics of Schizophrenia

Schizophrenia is a spectrum of disorders rather than a single disorder and can vary greatly in severity and specific symptoms between individuals[4]. The Diagnostics and Statistical Manual IV-R outlines that a clinical diagnosis of schizophrenia can be made if an individual has two or more of the following active symptoms that are present for a significant amount of time over the period of a month: delusions; hallucinations; disorganized speech; disorganized or catatonic behavior; and negative symptoms. If delusions are considered to be bizarre, or if hallucinations involve a voice keeping running commentary on a person's thoughts or behavior, or involve two or more voices conversing with each other, only one of the above symptoms need to be present for a diagnosis.

Negative symptoms are a loss of normal abilities and/or behavior. They usually contribute to poorer prognosis, and are generally more resistant to treatment. These can include:

- Alogia: inability to speak
- Affective flattening: loss or lack of emotional expressiveness, where one acts neither depressed nor happy
- Avolition: lack of motivation or drive, an individual is unable to pursue personally meaningful goals

Schizophrenia is divided into 5 clinical subtypes. An individual cannot be diagnosed with more than one subtype. Undifferentiated type is a diagnosis used when an individual does not meet the criteria for any of the other four subtypes.

Paranoid type: Characterized by frequent and persistent auditory hallucinations, or one or more delusions. Disorganized speech, disorganized behavior, and flat/inappropriate affect are minimal.

Disorganized type: Centered on disorganized speech and behavior, as well as flat/inappropriate affect.

Catatonic type: Major distinctive features include 2 or more of the following:

- Stupor/immobility
- Excessive and purposeless motor activity
- Echolalia or echopraxia (parroting words or movements of another person).
- Extreme negative symptoms with:

    - Resistance to following instructions, remaining rigid while others attempt to move you, or mutism
    - Voluntary movement peculiarities like noticeable grimacing, assumption of bizarre poses, stereotyped movements, and prominent mannerisms
- Abnormal voluntary movements:
    - Posturing (holding an unusual pose)
    - Noticeable grimacing
    - Stereotyped movements
    - Prominent mannerisms

Undifferentiated type: A subtype in which the individual meets the diagnostic criteria for schizophrenia, but the pattern of prominent symptoms is not characteristic of residual, paranoid, catatonic, or disorganized type.

Residual type: Defined by a lack of prominent delusions, hallucinations, catatonic behavior, disorganized behavior, or disorganized speech. Continued evidence of the underlying disorder is displayed in the form of negative symptoms, or 2 or more other symptoms with a milder phenotype (e.g. odd beliefs, unusual perceptual experiences).

Table 1. Prominent features of subtypes of schizophrenia.

	Schizophrenia Subtypes				
Possible Symptoms	Paranoid	Disorganized	Catatonic	Undifferentiated	Residual
Delusions	✓	—	—	—	—
Hallucinations	✓	—	—	—	—
Disorganized speech	—	✓	—	—	—
Disorganized/ catatonic behavior	—	✓	✓	—	—
Negative symptoms	—	✓	✓	—	✓

Impact on Quality of Life & Society

Suicide is a major risk factor, especially for young men[3]. It is a major contributory factor to the 15 year decrease in life expectancy among the population with schizophrenia[3]. Life span is further reduced by 10 years in the approximately 9-16% of the population with schizophrenia who will become homeless[3]. African Americans with schizophrenia have a greater risk of homelessness than other major ethnic groups in the United States, most of whom will have difficulty navigating the health system[3]. Disorganization and lack of income increases proportionally with disease severity; this is compounded with poor access to community housing and support services to create an environment where 48% of these individuals are not receiving care for mental health and the other chronic conditions they might have[3].

Genetic Counseling and Estimated Risk

Improved outcomes for people with schizophrenia can be achieved with early diagnosis, which can limit the severity and progression of disease[2,5], and through education of the genetic nature of the disease to help reduce stigma and improve societal and familial support for this vulnerable population[2,3,6,7]. Common population susceptibility alleles for schizophrenia are thought to be numerous, but the exact number and type of genetic variants is poorly understood[8]. It is thought that a subset of genes cause disease in individuals, so that any single susceptibility gene is neither necessary nor sufficient for causing disease. With heritability estimated to be 73-90%[2] the hope is to be able to identify possible high risk variants, or panels of small risk variants that put individuals at increased risk of developing schizophrenia[2,7,9]. In this way, the necessary social, familial, and medical supports and resources can be put in place before onset or early during onset of the disease to minimize negative outcomes[3,7].

This vulnerable and underserved population of individuals and their families need health professionals well-versed in technical, familial, and societal implications of schizophrenia[2,3]. Genetic counselors have long applied themselves to disease risk estimation, coordination of medical and social supports, and educational outreach[2]. Current psychiatric risk is evaluated using known disease associations and empirical evidence applied to the incidence and relatedness of schizophrenia and related neurodevelopmental disorders within a family[2]. Adding individual and/or family risk variant information has the potential to revolutionize the field of psychiatric genetics[2,7], as maternal serum screening has transformed the field of prenatal genetics.

Heritability

Schizophrenia is a complex disorder thought to involve multiple genes and the environment with heritability estimated between 73-90%[1,8,10]. About 60% of schizophrenia cases are sporadic with no affected first or second degree relatives [11]. Risk relative to the population increases with closer and increased numbers of affected relatives with up to 50% risk for individuals with an affected monozygotic twin[12]. Having a first degree relative poses an individual risk of 6% if one has a parent with schizophrenia, and up to 17% if one has a fraternal twin with schizophrenia. This large range of risk amongst individuals with the same relatedness to those with disease illustrate the environmental component to schizophrenia risk, which may include uterine environment, generational differences, social environment, and the family environment[13]. Although highly heritable, gene finding efforts for schizophrenia using family-based linkage analysis and the candidate gene approach have so far met with limited success[14].

Great hope was put in genome wide association studies (GWAS) to find risk variants for complex diseases like schizophrenia[9]. Focusing on single nucleotide polymorphisms (SNPs) researchers struggled to find the power to detect the many variants contributing to heritability that were thought to be found in low frequency and/or have a small effect on total disease risk[9,15]. Some of the limited number of loci with replicated results include: *ZNF804A*, which encodes a zinc finger transcription factor[16]; an upstream SNP of the *NRGN* that encodes a postsynaptic protein kinase[16]; and *TCF4*, a gene involved in cognitive function[16]. In recent years we have seen the culmination of multiple GWAS with increased power made possible by including large numbers in study and control groups, and the meta-analysis of these studies. Three separate groups have found associations in the major histocompatibility complex region on chromosome 6p21.3-22.1[16,17], which contains genes related to immunity, cognition, memory, and brain development[17]. The outcomes of these studies have helped build knowledge of the components of heritability of schizophrenia, but have only explained a small portion, as low as 6%[15], of the total estimated 80% heritability of schizophrenia.

More focus started to be put on copy number variants (CNVs) as an alternative for explaining the missing heritability in schizophrenia and related multifactorial neurodevelopmental diseases[9]. The typical number of copies of a genomic region is two, one maternal copy and one paternal copy. CNVs may be present as deletions where only one copy, or zero copies are present; or as duplications, where two or more copies of a genomic region are present. CNVs are one of the most common sources of variation making up 13% of the human genome, and an estimated 1000 CNVs in the average person[18]. CNVs have a wide range of physical size (hundreds – millions of base pairs) and may carry small or large risk for a specific disease or group of diseases. By studying risk effects of single CNVs and groups of CNVs,

genetic counselors and the medical community will gain an important tool to help improve and personalize risk estimates in order to facilitate early diagnosis and minimize negative disease outcomes.

Chapter 2: Real Time PCR Validation of Select CNVs in an Irish Population

Introduction

Copy Number Variants (CNVs) in the Human Genome

CNVs are localized duplication or deletion events that are thought to be caused mainly by nonallelic homologous recombination (NAHR), nonhomologous end-joining, or fork stalling and template switching (FoSTeS)[19,20]. NAHR is thought to be the most frequent cause of CNV formation and involves two regions of the genome that share low copy repeats (LCRs) that pair up and recombine during meiosis or mitosis[20]. If the two LCRs are in the same 5' to 3' orientation and are located on the same chromosome the result is a duplication and/or deletion in progeny cells. For efficient NAHR, LCRs must share minimal efficient processing segments (MEPS) – highly similar sequences that are typically 300 – 500bp, vary in length requirements depending on the involved loci, and can differ between meiotic and mitotic events. Certain regions of LCRs are genomic hotspots for NAHR, which explains the clustering of breakpoints and similar size of recurrent CNVs across multiple individuals. These regions often have features than can induce double-strand breaks, such as palindromes or minisatellites. Non-recurrent, rare CNVs can be mediated by NAHR through *Alu* and LINEs (long interspersed elements) and can result in clustering of breakpoints of CNVs. These rare CNVs may share one or more smallest region of overlap (SRO) across individuals, which can be used to investigate the association of specific SROs and any gene(s) contained within them with disease. In addition, some non-recurrent CNVs may have grouping of one breakpoint within a small genomic region across individuals, which may reflect the presence of genomic regions capable of double strand breaks.

Non-homologous end-joining (NHEJ) is normally used as one of the main mechanisms to repair double-strand DNA breaks but can also cause translocations in cancer, severe-combined

immunodeficiency through inherited NHEJ defects, or small CNVs[20]. After a double-strand break occurs and is detected, the break is bridged molecularly and the ends of the break are modified to make them compatible before repair of the break through ligation. Modification of the broken ends can involve the deletion or addition of several nucleotides, which has the potential to contribute to disease risk if located within a gene or another important genomic sequence.

FoSTeS is thought to be induced by repetitive sequences, palindrome, stem-loop structures, and other genomic features that cause DNA polymerase II to stall, which can cause the lagging replication strand to disengage and anneal with a new replication fork through microhomology[20]. Once DNA synthesis restarts, the invading strand primes its own template-based extension at the new fork. Multiple FoSTeS events can occur consecutively, which can result in complicated arrangements of sequences with a mixture of duplications and deletions in direct and/or inverted orientation to the original sequence.

Hypothetical Mechanisms of Disease

CNVs caused by the above mechanisms can encompass a portion or all of a gene, many genes, or no known genes[19]. While change in copy number can be relatively benign, change in copy number at certain loci in some individuals can cause or increase risk for disease through a number of different mechanisms such as gene position effects, gene disruption, gene dosage, and gene fusion events[19]. CNVs have been implicated in Mendelian diseases, such as DiGeorge syndrome, as well as complex multifactorial diseases like autism, intellectual disability, and schizophrenia[21].

Compared to controls, the overall number of CNVs has been shown to be increased in individuals with schizophrenia[22]. The International Schizophrenia Consortium found that rare CNVs that were greater than 100 kb and found in less than 1% of their sample were increased in number by 1.15-fold in individuals with schizophrenia compared to controls[8]. When taking genic content into account the difference in number of CNVs between cases and controls was 1.41-fold[8]. Affymetrix 5.0 SNP array genome wide scans of a study using trios of schizophrenia cases and controls showed an 8-fold increase of de novo CNVs in sporadic cases compared to controls[23]. Inherited CNVs were elevated by 1.5-fold in sporadic cases compared to controls[23]. This effect is more pronounced with CNVs that are both rare (<1%) and large (>1 Mb). These deletions and duplications were elevated 2.26-fold in cases vs. controls and up to 4.53-fold in cases when including only deletions[24].

CNV Risk Loci in Schizophrenia

A small number of rare, recurrent CNV loci have shown evidence of association with schizophrenia, including deletions at 1q21.1, 15q11.2, 15q13.3, and 22q11.2[8,17,22,24]. Other less-replicated CNV loci associated with schizophrenia risk include the 3q29 deletion, 16p11.2 duplication, 16p13.11 duplication, and 17q12 deletion[8,17,21,22,24]. Some of these CNVs have large effect sizes[25] including up to a 30% risk of schizophrenia for 22q11 deletion[21].

Common CNVs may also contribute to schizophrenia risk but are thought to have a smaller average effect size on disease risk compared to rare variants[21]. The current predominant hypothesis is that the mechanism for development of schizophrenia is most likely a combination of both rare and common CNVs, other genetic variants, and environmental factors[25].

Individuals with schizophrenia carrying the most-studied risk conferring CNV, the 22q11.2 deletion, are clinically indistinguishable in terms of schizophrenia diagnosis from those who do not carry the deletion. The frequency of the 22q11.2 deletion in the general population is about 1/3,000 - 1/6,000 live births[6,13], while the frequency in the population with schizophrenia is about 1%[13]. On average, approximately 90-95% of deletions are de novo, and 5-10% are inherited, with affected family members showing variable expressivity and differing clinical features[6,21,26,27]. In clinical settings transmission is often seen through mothers with mild neuropsychiatric phenotypes[6]. Risk of schizophrenia with a 22q11 deletion is about 25-30%, and individuals with 22q11 deletion account for about 1% of the population with schizophrenia making the 22q11 deletion the variant with the largest risk effect size discovered to date for schizophrenia.

The 22q11 deletion can cause highly variable phenotypes. Common features include cleft lip/cleft palate, intellectual disability, developmental delay, seizure, immunodeficiency, and congenital heart defects[13,26,27]. Most cases discovered to date have similar microdeletions 3 Mb in length and the high recurrence rate of these mutations may be related to low-copy repeats that predispose this region to rearrangements[13]. Most of the deletions seen at this locus are thought to originate from non-allelic homologous recombination events[6]. Genetic background and environmental factors likely play a role in the observed variable phenotypes in individuals with similar or even identical deletions.

CNV Detection and Validation

The three main methods of CNV detection are comparative genomic hybridization, using raw SNP intensities from GWAS SNP array, and direct sequencing[28-30]. Comparative genomic

hybridization was long considered to be the gold standard for CNV detection, but there is significant cost involved. Using algorithms to call raw CNVs from GWAS SNP arrays is being increasingly used in research settings, because it is currently the most cost effective method and because SNP intensity data is often freely available from previous GWAS studies[28,29]. This study aims to explore feasibility of calling CNVs from SNP intensity data in terms of accuracy and validity of calls, and whether follow-up validation with real-time qPCR is necessary.

CNV Detection Algorithms using Single Nucleotide Polymorphism DNA Microarray

CNV detection algorithms attempt to determine locations of CNVs in an individual's genome by looking for changes in sets of sequential SNP fluorescent probe intensities along a chromosome, often 10 probes or more in length[28-30]. A putative CNV region will have a decreased probe intensity for deletions (approximately 50% decrease for single copy, very low or non-detectable signal intensity for zero copy), and increased probe intensity for duplications (150% for three copies), compared to the surrounding probe intensities in the putative non-CNV (diploid) region. Duplications greater than three copies in number become more difficult to ascertain exact copy number because the difference in the increase in probe intensity becomes smaller with increasing copy numbers.

SNP probe intensities vary according to genomic location, quality of the probe design, and between technical and biological replicates, which can result in a higher false positive rate and possibly a higher false negative rate compared to the other methods of CNV detection[28-30]. A number of studies have tried to address specificity and sensitivity of CNV calls using SNP intensities by using multiple calling algorithms. While this may reduce the false positive rate, the sensitivity may be severely reduced. Specific algorithms like PennCNV are thought to call CNVs

much more conservatively than other algorithms, which would reduce the amount of overlap seen between multiple algorithms[30]. In certain cases, such as CNVs less than 10 kb in size it may be preferable to require the agreement between two or more algorithms before making a CNV call.

Although some progress has been made in improving CNV calling through these algorithms it is still currently necessary to follow up putative CNV loci with validation through a method like quantitative real-time PCR, especially if a particular locus has not been previously identified in the literature. This study analyzed CNVs from an Irish case control and family sample to further evaluate the utility of CNV-based risk analysis for schizophrenia using microarray SNP intensity algorithm calls.

Some of the CNVs evaluated in this study that have prior evidence for association with schizophrenia include regions encompassing *chromodomain helicase DNA binding protein 1-like* (*CHD1L), p21 protein (Cdc42/Rac)-activated kinase 7 (PAK7),* and *v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4 (ERBB4). CHD1L* is located at 1q21.1, where changes in copy number have been associated with schizophrenia, intellectual disability, autism, dysmorphic features, congenital heart anomalies, or a normal phenotype with no detectable clinical abnormalities[31]. Both duplications and deletions are present in this region that includes at least 12 genes, including *CHD1L*. Harvard *et. al.*[31] showed a correlation between RNA and protein expression and CNVs at *CHD1L*. Reduced *CHD1L* in lymphoblast cell lines was shown to interfere with a cell cycle checkpoint, and resulted in elevated levels of pseudomitotic cells that exhibited entangled chromatids. This phenotype is similar to that seen in Werner syndrome, which predisposes individuals to premature aging and cancer.

Although not statistically significant, two studies found *PAK7* duplications in schizophrenia cases only and not in controls[32]. *PAK7* is a Ser/Thr protein kinase expressed predominantly in the brain, which is implicated in regulation of cell proliferation, survival, and signaling, as well as cytoskeletal dynamics[33]. It may provide a role in neurite development by promoting neurite outgrowth.

A common CNV has been observed in intron 1 of *ERBB4*. ERBB4 is a receptor for neuregulin, a ligand potentially involved in glutamatergic synapse plasticity and NMDA receptor hypofunction in schizophrenia[34] . This is hypothesized to influence the onset of positive symptoms like hallucinations and may be precipitated by stress. A number of candidate gene and association studies have provided support for an association between neuregulin and schizophrenia risk, and *ERBB4* has been linked with increased risk for childhood schizophrenia[32,35].

Cytochrome c oxidase subunit Vb (COX5B), and *zinc finger, FYVE domain containing 20 (ZFYVE20)* genes have been implicated in psychiatric and neurodevelopmental disorders such as depression and autism. Rat studies looking at the oxidative phosphorylation pathway showed that *COX5B* may provide a protective effect against stress. Rats resilient to experimental stressors had significantly up-regulated *COX5B* expression. It has been hypothesized that susceptibility to stress plays a major role in the development of unipolar depression and post-traumatic stress syndrome, and some studies have linked stress susceptibility to development of schizophrenia as well[36-38].

Materials and Methods

Study Populations

CNVs were studied using the Irish Case Control Study of Schizophrenia sample (ICCSS) (n=1021 cases)[39] and Trinity Biobank Irish controls (n=2000 controls). A common CNV in *ERBB4* was evaluated using the aforementioned case and control samples, as well as the Irish Sample of High Density Schizophrenia Families (ISHDSF) (n=270 families, n=1426 individuals). Affected status in both ICCSS and ISHDSF was established using DSM-IIIR criteria[39] from in-patient and out-patient facilities in Northern Ireland and Ireland. ISHDSF were ascertained through Irish probands that had one or more additional family member affected with schizophrenia, or schizoaffective disorder. Data from North Americans of European descent were obtained from the Children's Hospital of Philadelphia (CHOP) by Shaikh, Gai, Perin, *et. al*[40].

CNV Calling from SNP Arrays

Putative CNVs were called in n=1021 ICCSS cases and n=2000 Trinity Biobank controls run on Affymetrix version 6.0, 900k SNP array from SNP intensities using Birdseye in Birdsuite (v 1.5.5). ISHDSF CNV calls were generated from PennCNV from Illumina 610 Quad array SNP intensities using select Irish Sample of High Density Schizophrenia Families (ISHDSF) samples that were independent from one another (family founders), n=107 [41].

Samples were excluded that had failed SNP quality control and where SNP call rate was <0.95. Calls were excluded from sex chromosomes; where putative CNV lengths were >10Mbp or <100kbp; and where the putative CNV had ≥50% overlap with a region that had CNVs in ≥1% of samples.

Putative CNV Locus Selection for Validation/Evaluation

The top CNVs from ICCSS and Trinity Biobank called by the Wellcome Trust, as chosen by locus significance and recurrent CNV loci, were selected for validation at VCU and Trinity College Dublin (TCD) through real-time quantitative PCR. Regions selected using genome build hg18[42] include: chromosome 2cen – q13, 97-98Mb (*COX5B)*; chromosome 1q12, 144 – 147Mb (*CHD1L*); chromosome 3p25.1, 15.05 – 15.2Mb (*ZFYVE20*); and chromosome 20p12, 9 – 10Mb (*PAK7*). A putative common CNV at chromosome 2, 211 – 213Mb (*ERBB4*) was selected based on previous association with schizophrenia, and a possible over-representation in ISHDSF family founders compared to CHOP North American controls in the literature (n=2640)[40]. *ERBB4* deletion frequency was also assessed in ICCSS (n=462) and Trinity Biobank samples (n=448) using real-time qPCR.

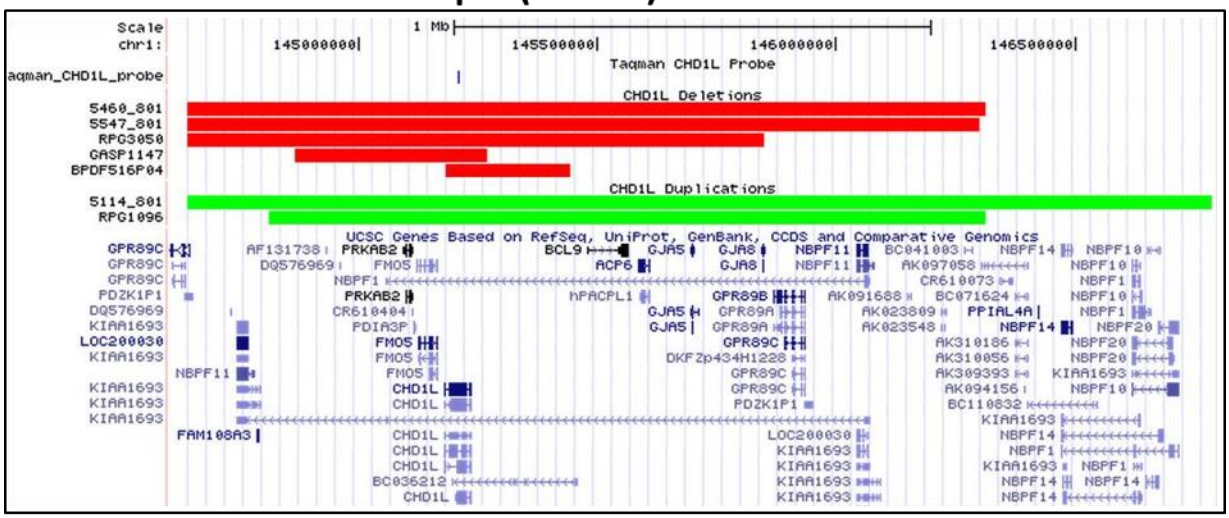
Real-Time Quantitative PCR Validation of Putative CNVs

Invitrogen Taqman probes were chosen using the smallest region of overlap shared between all samples within a putative CNV region for which a pre-designed Taqman probe was available (Figure 1). All DNA samples for cases and controls were quantified using Nanodrop (Thermo Scientific) and transferred to Bio-Rad 200μL 96-well plates, where each sample concentration was adjusted to within 50 – 55 ng/μL using DNA sample buffer. Working plates of 5 – 6 ng/μL DNA were made using the Eppendorf epMotion 5075 robot on ABI Prism 384-well clear optical reaction plates. One blank (DNA suspension buffer) and one CEPH control sample (Coriell Institute for Medical Research) were included on each working plate.

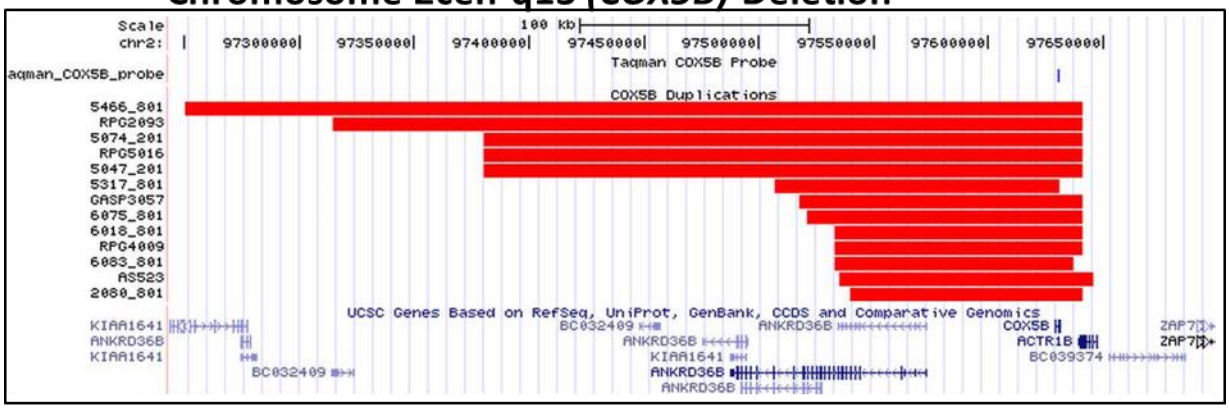
PCR reactions are set up using the Eppendorf epMotion 5075 robot with 50 μL Eppendorf epTIPS Motion pipette tips and 2 μL of 5 – 6 ng/μL DNA and 8 μL of PCR reaction mix containing ABI Taqman primer mix for the specific gene being tested, ABI RNase P Taqman primer mix (internal standardization control), Taqman reaction mix, and DNase and RNase free distilled water. Quadruplicate samples were then run using templates called for CNVs on the ABI 7900HT fast real-time PCR system machine.

Quantitative real-time PCR reactions were analyzed using SDS software (Applied Biosystems) with a CT threshold of 0.2, and quantification data from SDS was exported as a tab-delimited text file and imported to Copy Caller in this format. For each sample, Copy Caller determined a confidence interval for copy number and then assigned a best full integer estimate for copy number. Copy number calls were normalized to the average of the most common copy number variant present at the locus in the samples analyzed. Samples with copy number confidence scores below 0.84 were excluded from further analysis. ISHDSF data was provided courtesy of Erik Loken.

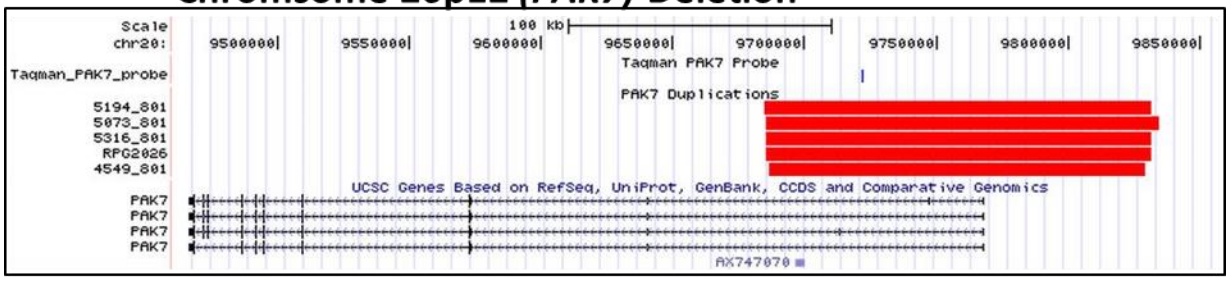
Chromosome 1q12 (CHD1L) Deletion



Chromosome 2cen-q13 *(COX5B)* Deletion



Chromsome 20p12 *(PAK7)* Deletion



Chromosome 3p25.1 *(ZFYVE20)* Deletion

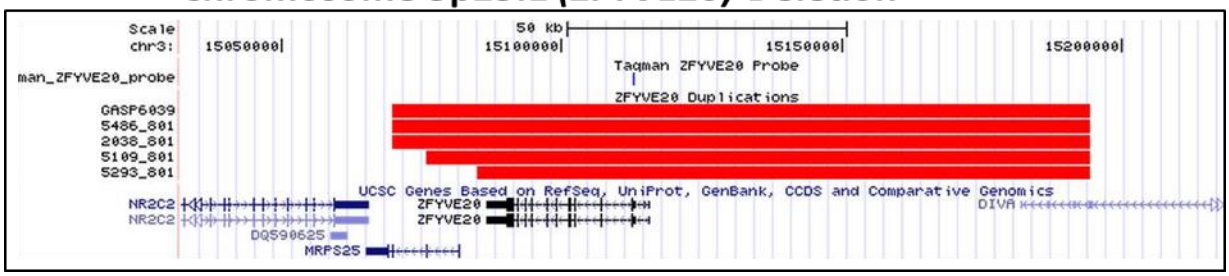


Figure 1. UCSC genome browser custom track (hg18 assembly) showing Taqman probe location for qPCR validation of ICCSS samples with putative CNVs. Red bars: subjects with putative deletion CNVs. Green bars: subjects with putative duplication CNVs. Known UCSC genes are shown below putative CNVs (http://genome.ucsc.edu)[42].

Results

ISHDSF and ICCSS sample validation of common copy number variant in *ERBB4*

A common deletion variant in *ERBB4* was found to have an elevated frequency in the ISHDSF sample using PennCNV on SNP intensities from microarray data on select samples (Tables 2-3). PennCNV calls were validated using Taqman quantitative real-time PCR and showed that the frequency in independent family founders was 9.8% (n=107) (courtesy Erik Loken), which was increased compared to 4.5% in North American Caucasians (Children's Hospital of Philadelphia sample) (chi-sq=12.2, p=0.0005) (Table 2). The ICCSS case sample and the Trinity Biobank Irish control sample were used to rule out the possibility of population-specific frequencies of the *ERBB4* deletion variant. Table 3 summarizes the *ERBB4* deletion frequencies found in the samples used. Taqman qPCR validation showed that there is an increase in deletion variant frequency in Irish controls (7.7%, n=448) compared to North American Caucasians (4.5%). No significance difference was found when comparing the deletion frequency in Irish Family Founders to population matched controls (chi-sq=1.0, p=0.31), or Irish cases (6.6%, n=462) to population controls (chi-sq=0.8, p=0.36).

Table 2. Real-time qPCR assessment of the *ERBB4* CNV.

Predicted Copy Number	ICCSS Cases	Trinity Biobank Controls	IHDS Family Founders**
0	1	2	0
1	59	65	21
2	400	381	86
3	1	0	0
4	0	0	0
5	1	0	0
Total Analyzed*	462	448	107

*Calculated copy number calls with confidence scores <0.84 were excluded. **IHDS Family data courtesy Erik Loken.

Table 3. *ERBB4* deletion frequencies in Irish family and singletons with schizophrenia compared to North American and Irish controls.

Sample	Deletions (%)*	Non-Deletions
Independent Family[41]**	21 (9.8%)	193
North American Caucasian[40]	118 (4.5%)	2522
Irish Singleton Cases	61 (6.6%)	863
Irish Population Controls[39]	69 (7.7%)	827

Irish Cases vs. Irish Controls $\chi^2_{df=1}$= 0.032, p=0.858. Family vs. Irish Controls $\chi^2_{df=1}$ =0.321, p=0.571. *Calculated copy number calls with confidence scores <0.84 were excluded. **IHDS Family data courtesy Erik Loken.

ICCSS sample validation for rare copy number variants

Four rare CNVs called by the Wellcome Trust Consortium[39] were validated using Taqman real-time qPCR in 29 samples. Each sample contained a single CNV in one of the loci evaluated: *CHD1L* (1q12), *COX5B* (2cen-q13), *PAK7* (20p12), *ZFYVE20* (3p25.1) (Table 4, Figure 2). The Birdseye in Birdsuite algorithm called 1 duplication at *COX5B* and no CNVs at the other 3 loci in Irish Biobank controls. Birdseye called 7 CNVs at *CHD1L*, 12 CNVs at *COX5B*, and 5 CNVs each at *ZFYVE20* and *PAK7* loci. All four loci showed locus-specific higher rates of CNVs in cases compared to controls, but none of these loci except for *COX5B* achieved genome-wide significance (Table 4).

These samples were concurrently validated, using separate qPCR assays, by the Dublin group (Table 6). Validations performed at VCU and Dublin were in agreement except for two samples, and each of the four CNVs was validated at the selected loci. The calls not in agreement were: a *COX5B* duplication found in a Biobank control sample validated at VCU, C0822, was not validated by Dublin; and case sample 5301-801, where the VCU group found a duplication and the Dublin group found no CNV. VCU and Dublin found a *CHD1L* deletion in one sample where Birdseye found no CNV.

Table 4. Potential CNV risk loci for schizophrenia validated by qPCR in an Irish case-control sample.

Name	Symbol	Position*		CNV Size	# Cases	# Controls	Broad** Dup/Del Significance Level
Chromodomain helicase DNA binding protein 1-like	*CHD1L*	1q12	146.7 Mb	258kb – 2143kb	7	0	Locus P: 0.0125399 GW P: 0.446036
Cytochrome c oxidase subunit Vb	*COX5B*	2cen-q13	98.3 Mb	101kb – 391kb	12	1	Locus P: 0.000299997 GW P: 0.00805992
Zinc finger, FYVE domain containing 20	*ZFYVE20*	3p25.1	15.1 Mb	110kb – 125kb	5	0	Locus P: 0.0444796 GW P: 0.943271
p21 protein (Cdc42/Rac)-activated kinase 7	*PAK7*	20p12	9.5 Mb	142kb – 149kb	5	0	Locus P: 0.0433896 GW P: 0.943271

*Genomic positions based on hg18 assembly. ** Broad schizophrenia spectrum: includes all disorders that significantly aggregated in relatives of schizophrenic probands in the Roscommon Family Study[43]

CHD1L

Sample	Birdseye	VCU	Dublin
GASP1147	del	del	del
BPDF516P04	del	del	del
RPG1096	dup	NT	dup
RPG3050	no CNV	del	del
5460-801	del	del	del
5547-801	del	del	del
5114-801	dup	dup	dup

ZFYVE20

Sample	Birdseye	VCU	Dublin
GASP6039	dup	dup	dup
GASP3046	dup	NT	dup
2038-801	dup	dup	dup
5293-801	dup	dup	dup
5486-801	dup	dup	dup
5109-801	dup	dup	dup

COX5B

Sample	Birdseye	VCU	Dublin
RPG5016	dup	dup	dup
GASP3057	dup	dup	dup
RPG4009	dup	dup	dup
C0822	dup	dup	no CNV
RPG2093	no CNV	NT	dup
2080-801	dup	dup	dup
5074-201	dup	dup	dup
5317-801	dup	dup	dup
5466-801	dup	dup	dup
6018-801	dup	dup	dup
6075-801	dup	dup	dup
6083-801	dup	dup	dup
5047-201	dup	dup	dup
5301-801	dup	dup	no CNV
AS523	no CNV	NT	dup

PAK7

Sample	Birdseye	VCU	Dublin
RPG2026	dup	dup	dup
5073-801	dup	dup	dup
5194-801	dup	dup	dup
4549-801	dup	dup	dup
5316-801	dup	dup	dup

Figure 2. CHD1L, COX5B, PAK7, and ZFYVE20 CNV validation of Birdseye calls by VCU and Dublin groups in ICCS sample using Taqman qPCR**.** Rare CNVs were called from Affymetrix 6.0 SNP intensities using Birdseye in Birdsuite CNV calling algorithm. Green: CNV duplications. Red: CNV deletions. Yellow: No discernible CNV. White: Sample not tested*. *Specific samples were unavailable for testing at VCU due to low amount of DNA available.

Discussion

Four rare CNVs called by the Wellcome Trust Consortium using Birdseye in Birdsuite with Affymetrix 6.0 array raw SNP intensities, *CHD1L*, *COX5B*, *PAK7*, *ZFYVE20*, were validated using distinct Taqman real-time qPCR assays in 29 samples at VCU and Dublin, and were in agreement except for two samples. The lack of agreement could be because of the difference in Taqman probes used between the two groups: the probes may have different properties affecting PCR efficiency that may have made the select Taqman qPCR assays used by the Dublin group less sensitive to detecting duplications than the comparable qPCR assays used by the VCU group. In addition, there were technical issues regarding the age and quality of DNA samples that required repeated and careful titration to ensure the DNA of the samples and controls used were of near equal concentrations. Detection of duplications can be more sensitive to small deviations in DNA concentrations, since the goal is to detect a 50% difference in copy number compared to diploid controls. Detection of deletions can tolerate a greater deviation in DNA concentration amongst samples, since there is a two-fold difference in copy number compared to diploid controls; all of the putative deletion CNVs were in agreement between the VCU and the Dublin group. Validation could be repeated at both sites using both sets of Taqman assays for each locus, which would show whether the discrepancies in calls were the result of differing Taqman assay sensitivities, or because of some other technical or analytical difference between the two sites.

VCU and Dublin called a *CHD1L* deletion that Birdseye called as regular copy number (diploid). This appears to be a false negative, since qPCR is a more robust method of CNV detection, and because the call was in agreement at two separate sites performed with two separate qPCR assays. False negative calls such as these cause a reduction in analytical power

that can make it more difficult to achieve genome-wide significance for true differences in frequencies between affecteds and controls. An additional difficulty is power reduction due to the small effect sizes typically seen with CNVs associated with schizophrenia risk. These two issues, along with lack of consistency of CNV validation across test sites and/or across test assays present obstacles to wide-spread clinical use of schizophrenia risk screening using CNVs called from microarray SNP intensities.

Although all four of these loci showed locus-specific significance, none of the loci achieved genome-wide significance for difference between cases and controls except for *COX5B* (Table 4). Low effect size and rarity of CNVs at specific loci decrease analytical power for finding associations between loci and schizophrenia risk. Meta-analysis of multiple studies may be necessary to achieve genome-wide significance for many schizophrenia-risk CNV loci, since many of the loci with hypothetical risk discovered to date tend to be rare[44].

One common CNV, *ERBB4*, was evaluated by real-time qPCR in both ICCSS and IHDSF samples and CNV deletion frequencies were compared with Trinity Biobank and North American control populations. The *ERBB4* deletion frequency was significantly different when comparing the Irish family to the North American controls. The deletion frequency was not significantly different when comparing the Irish controls to the Irish singleton or the Irish family samples. Because the deletion frequency did not differ when population-matched controls were compared to family and case samples this suggests that population stratification was the cause of the difference seen between Irish family and North American controls, rather than a true association of the deletion with schizophrenia.

The Irish population was chosen as a good candidate for study of schizophrenia due to the low immigration and emigration rates compared to North Americans. All affecteds were

screened for Irish ethnicity of both maternal and paternal grandparents. CNVs at specific loci, like *ERBB4*, can differ in frequency between different homogeneous populations, such as the population found in Ireland, or when compared to heterogeneous populations like North Americans, because of founder effects and/or random genetic drift. This study illustrates the necessity of comparing affected samples to appropriately matched ethnic controls.

Future Directions

Once associations between a CNV locus and schizophrenia risk are shown based on difference in frequency between cases and controls, the next step is to discover whether that CNV affects expression of a specific gene(s). Although validation in this study was based on Taqman probes located within one specific gene, it is possible that the true disease association is based on a neighboring gene located within the same CNV. In addition, association with schizophrenia risk of gene(s) within a CNV may not necessarily be caused by a difference in expression of gene(s) within the CNV itself. One or more regulatory elements within a CNV may be affecting a gene(s) at another locus either intra- or inter-chromosomally. Alternatively a CNV could influence chromatin structure and/or organization of chromatin within the nucleus, affecting many genes at once.

To test for expression-specific effects on schizophrenia risk within CNVs in this Irish study one could make use of the lymphoblast cell lines that were established for many of the samples that were evaluated in this study. mRNA from lymphoblasts could be collected for use with real-time qPCR to look for difference in mRNA expression. Cases with and without the specific CNV could be compared for mRNA expression differences to see whether the CNV itself has a direct effect on expression. Cases with and without the CNV could also be compared

to controls with and without the CNV to see whether there is a difference in gene expression between cases and controls that occurs regardless of the presence or absence of the CNV. This latter experiment could shed light on the overall importance of a specific gene's expression in schizophrenia pathogenesis. It is possible that a gene's expression could be influenced directly by the presence of a CNV, or indirectly, by epigenetics or gene-gene interactions that may be shared in common across of subset of individuals with schizophrenia.

Any genes with mRNA expression differences should also be followed up with western blot to look for difference in lymphoblast protein expression. This is important because proteins are much closer to the final phenotype output that may or may not influence schizophrenia risk, and many mRNA changes do not correlate directly with changes in protein levels. Western blot may not be available for all loci due to possible scarcity of adequate locus-specific antibodies; risk for loci with only mRNA expression data available should be viewed as weaker evidence compared to those with protein expression data.

Chapter 3: Discussion

Current Clinical Practice of Genetic Counseling for Schizophrenia

Genetic counseling for schizophrenia risk is currently based on empirical risk estimates based on family history (Figure 1), as well as counseling specific for the few rare recurrent high-risk CNVs identified to date: 1q21.1, 15q13.3, and 22q11 deletions[21,45]. The utility of empirical risk is limited somewhat in terms of providing specific risk with the goal of providing anticipatory care: it is limited by availability of data for only relatively simple family structure with few affected relatives; multiple types of psychiatric disorders can aggregate in families, which can't be used to improve risk estimates with empirical data currently available; and the range of risk provided is often large leaving patients with a degree of uncertainty that can possibly instill a false sense of security, or conversely, instill anxiety from perceived heightened risk of developing schizophrenia[45]. Despite these difficulties a positive family history remains the single greatest risk factor for developing schizophrenia.

Application of CNV-related Risk Estimates in Clinical Practice

Currently known CNVs may not impact risk management as much as family history, but the clinical utility cannot be ignored. High risk CNVs for schizophrenia are rare and are often inherited *de novo*, but they have a predictable manner of transmission that can be used to provide individualized and more exact risk estimates for children of a known carrier[21,45,46]. Chance of transmission to offspring is 50%, although parents should be made aware that the expressivity of disease, including schizophrenia, can vary greatly even between family members with identical variants[21].

Specific penetrance for known recurrent high-risk CNVs vary from as high as 55% for 22q11.2 deletion, to as low as 2% for 2p16.3 deletion and 15q11.2 deletion[46]. Schizophrenia risk from these CNVs is modified by other CNVs, and additional genetic and environmental factors; none of these individual CNVs are necessary, or sufficient for disease. The rarity coupled with the penetrance of these CNVs makes routine clinical screening for these variants unnecessary and not cost effective in the absence of additional clinical features. Clinical screening would be suggested if personal or family history is suggestive of features associated with a chromosomal disorder that could indicate presence of a pathogenic CNV such as: presence of facial dysmorphology, one or more birth defects, intellectual disability, a history of unexplained seizures, or another feature in the proband suggestive of a chromosome condition[21]. The most promising current application may be for individuals with a known schizophrenia diagnosis: previously unidentified pathogenic CNVs have been identified in approximately 2.5% of these individuals[21,47]. This could impact medical management outside of their schizophrenia diagnosis, as these syndromes are often associated with a number of other phenotypes, some of which may be potentially fatal, and/or not easily detectible by routine clinical examination[48]. An example is previously undiagnosed congenital heart conditions that can be associated with 22q11.2 deletion. About 1% of individuals with schizophrenia also have been shown to have a previously undiagnosed 22q11.2 deletion, which can have a major implication on their health outside of their psychiatric disorder[47].

Future of Psychiatric Genetic Counseling

While individual CNVs may be limited in terms of providing population risk estimates, gene pathway and network analysis using CNVs associated with schizophrenia may hold promise in expanding our current knowledge of schizophrenia etiology and treatments[2,6,45]. Multiple rare CNVs can be grouped by common function or pathways in order to discover associations with disease that would be unlikely to be found individually due to small effect size and small numbers of individuals affected with the single CNVs[35,49]. Functional genetic groups may be used to discover endophenotypes that can predict disease status at an earlier age; biomarkers that can identify affected individuals more quickly and easily than current clinical diagnosis; or previously unidentified pathogenic CNVs[16,50].

Many clinicians are looking towards opportunities for improving the quality of life for individuals affected by schizophrenia by identifying at-risk individuals before onset, or early after onset of disease[6,13]. Pre-emptive care can be provided by establishing personalized neuro-imaging or neuro-cognitive precursors to psychosis; anticipatory care and/or management of symptoms in addition to psychosis that may accompany specific risk variants; developing tailored risk-estimates to aid in family planning of currently unaffected individuals; and attempting to limit environment or gene by environment interactions that further increase risk of disease. Known environmental contributors to risk have shown only small-modest effect, but can still be useful to minimize risk as much as possible. Factors suggested to be monitored for high-risk individuals include: substance use avoidance (particularly marijuana use early in life); physical and mental exercise, and good nutrition.

Cognitive and antipsychotic medication-based treatment of individuals has been shown to minimize disease severity and improve quality of life outcomes when implemented early after

onset of disease[6]. Aggressive treatment of early signs of schizophrenia can help minimize hospitalizations, and disruptions to employment, academic pursuits, and family and social interactions.

Summary

Ongoing evaluation of rare and common CNVs affecting *ERBB4I*, *ZFYVE20*, *COX5B*, *PAK7*, and *CHD1L*, is an important step towards identifying and confirming genetic risk factors for schizophrenia. Although utility of screening for similar variants in the clinical population remains limited, useful insights have been gained from study of rare, higher-risk variants, such as 22q11 deletion. CNV calling algorithms have improved tremendously with their continued use, but validation using real-time qPCR, or another similar method is still needed to confirm identification of CNVs. The majority of CNVs identified through calling algorithms did validate with qPCR in this study, but at present the reliability and accuracy of these calls is not high enough to be used in a clinical setting.

Future work will see continued exploration of use of multiple CNV-calling algorithms and adjustment of calling parameters to optimize accuracy and minimize false negatives[28,30]. This may eventually improve CNV calling enough to be able to use this method in a clinical setting. Benefits of this method could be quicker turn-around of patient results if the patient has had a previous SNP microarray performed, reduced cost of patient care, and potential screening of a broader patient population for high risk psychosis-related or other pathogenic CNVs.

High-risk individuals for schizophrenia, and other neurodevelopmental disorders, will hopefully see continually-increasing benefits from anticipatory care of mental health and other health conditions that may accompany specific genetic variants[13,47,51]. They may also look

forward to delayed disease onset and/or minimized severity of disease due to earlier diagnosis and treatment of disease.

List of References

References

1. Girard SL, Xiong L, Dion PA, Rouleau GA. Where are the missing pieces of the schizophrenia genetics puzzle? *Curr Opin Genet Dev*. 2011.

2. Collier DA, Vassos E, Holden S, Patch C, McGuire P, Lewis C. Advances in the genetics of schizophrenia: Will high-risk copy number variants be useful in clinical genetics or diagnostics? *F1000 Med Rep*. 2009;1:61.

3. Foster A, Gable J, Buckley J. Homelessness in schizophrenia. *Psychiatr Clin North Am*. 2012;35(3):717-734.

4. Braff DL, Ryan J, Rissling AJ, Carpenter WT. Lack of use in the literature from the last 20 years supports dropping traditional schizophrenia subtypes from DSM-5 and ICD-11. *Schizophr Bull*. 2013.

5. For improved outcomes, connect first-episode schizophrenia patients to comprehensive care quickly. *ED Manag*. 2012;24(8):93-95.

6. Costain G, Esplen MJ, Toner B, Hodgkinson KA, Bassett AS. Evaluating genetic counseling for family members of individuals with schizophrenia in the molecular age. *Schizophr Bull*. 2012.

7. Hill MK, Sahhar M. Genetic counselling for psychiatric disorders. *Med J Aust*. 2006;185(9):507-510.

8. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455(7210):237-241.

9. Gershon ES. After GWAS: Searching for genetic risk for schizophrenia and bipolar disorder. *Am J Psychiatry*. 2011;168(3):253-256.

10. Williams HJ, Owen MJ, O'Donovan MC. Schizophrenia genetics: New insights from new approaches. *Br Med Bull*. 2009;91:61-74.

11. Xu B, Woodroffe A, Rodriguez-Murillo L, et al. Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. *Proc Natl Acad Sci U S A*. 2009;106(39):16746-16751.

12. Kukshal P. Genetics of schizophrenia from a clinicial perspective. *International review of psychiatry*. 2012;24(5):393-404.

13. Bassett AS, Marshall CR, Lionel AC, Chow EW, Scherer SW. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Hum Mol Genet*. 2008;17(24):4045-4053.

14. Sullivan PF. Schizophrenia genetics: The search for a hard lead. *Curr Opin Psychiatry*. 2008;21(2):157-160.

15. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: A survey of ten complex diseases. *Genet Epidemiol*. 2011;35(5):310-317.

16. Haraldsson HM, Ettinger U, Sigurdsson E. Developments in schizophrenia genetics: From linkage to microchips, deletions and duplications. *Nord J Psychiatry*. 2011;65(2):82-88.

17. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460(7256):744-747.

18. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Prog Neurobiol*. 2012;99(1):81-91.

19. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451-481.

20. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics*. 2008;1(1):4.

21. Bassett AS, Scherer SW, Brzustowicz LM. Copy number variations in schizophrenia: Critical review and new perspectives on concepts of genetics and disease. *Am J Psychiatry*. 2010;167(8):899-914.

22. Grozeva D, Kirov G, Ivanov D, et al. Rare copy number variants: A point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch Gen Psychiatry*. 2010;67(4):318-327.

23. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008;40(7):880-885.

24. Kirov G, Grozeva D, Norton N, et al. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet*. 2009;18(8):1497-1503.

25. Doherty JL. Recent genomic advances in schizophrenia. *Clin Genet*. 2012;81(2):103-109.

26. Bassett AS, Chow EW. 22q11 deletion syndrome: A genetic subtype of schizophrenia. *Biol Psychiatry*. 1999;46(7):882-891.

27. Goldmuntz E. DiGeorge syndrome: New insights. *Clin Perinatol*. 2005;32(4):963-78, ix-x.

28. Zhang D, Qian Y, Akula N, et al. Accuracy of CNV detection from GWAS data. *PLoS One*. 2011;6(1):e14511.

29. Nord AS, Lee M, King MC, Walsh T. Accurate and exact CNV identification from targeted high-throughput sequence data. *BMC Genomics*. 2011;12:184.

30. Lin P, Hartz SM, Wang JC, et al. Copy number variation accuracy in genome-wide association studies. *Hum Hered*. 2011;71(3):141-147.

31. Harvard C, Strong E, Mercier E, et al. Understanding the impact of 1q21.1 copy number variant. *Orphanet J Rare Dis*. 2011;6:54-1172-6-54.

32. Need AC, Ge D, Weale ME, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet*. 2009;5(2):e1000373.

33. Pandey A, Dan I, Kristiansen TZ, et al. Cloning and characterization of PAK5, a novel member of mammalian p21-activated kinase-II subfamily that is predominantly expressed in brain. *Oncogene*. 2002;21(24):3939-3948.

34. Bennett M. Positive and negative symptoms in schizophrenia: The NMDA receptor hypofunction hypothesis, neuregulin/ErbB4 and synapse regression. *Aust N Z J Psychiatry*. 2009;43(8):711-721.

35. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320(5875):539-543.

36. Henningsen K, Palmfeldt J, Christiansen S, et al. Candidate hippocampal biomarkers of susceptibility and resilience to stress in a rat model of depression. *Mol Cell Proteomics*. 2012;11(7):M111.016428.

37. Fatemi SH. Reelin, a marker of stress resilience in depression and psychosis. *Neuropsychopharmacology*. 2011;36(12):2371-2372.

38. Pruessner M, Iyer SN, Faridi K, Joober R, Malla AK. Stress and protective factors in individuals at ultra-high risk for psychosis, first episode psychosis and healthy controls. *Schizophr Res*. 2011;129(1):29-35.

39. Riley B, Thiselton D, Maher BS, et al. Replication of association between schizophrenia and ZNF804A in the irish case-control study of schizophrenia sample. *Mol Psychiatry*. 2010;15(1):29-37.

40. Shaikh TH, Gai X, Perin JC, et al. High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications. *Genome Res*. 2009;19(9):1682-1690.

41. Hawi Z, Myakishev MV, Straub RE, et al. No association or linkage between the 5-HT2a/T102C polymorphism and schizophrenia in irish families. *Am J Med Genet*. 1997;74(4):370-373.

42. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.

43. Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The roscommon family study. I. methods, diagnosis of probands, and risk of schizophrenia in relatives. *Arch Gen Psychiatry*. 1993;50(7):527-540.

44. Hosak L, Silhan P, Hosakova J. Genomic copy number variations: A breakthrough in our knowledge on schizophrenia etiology? *Neuro Endocrinol Lett*. 2012;33(2):183-190.

45. Austin JC, Peay HL. Applications and limitations of empiric data in provision of recurrence risks for schizophrenia: A practical review for healthcare professionals providing clinical psychiatric genetics consultations. *Clin Genet*. 2006;70(3):177-187.

46. Vassos E, Collier DA, Holden S, et al. Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet*. 2010;19(17):3477-3481.

47. Bassett AS, Costain G, Fung WL, et al. Clinically detectable copy number variations in a canadian catchment population of schizophrenia. *J Psychiatr Res*. 2010;44(15):1005-1009.

48. Moreno-De-Luca D, Cubells JF. Copy number variants: A new molecular frontier in clinical psychiatry. *Curr Psychiatry Rep*. 2011;13(2):129-137.

49. Saus E, Brunet A, Armengol L, et al. Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *J Psychiatr Res*. 2010;44(14):971-978.

50. Chan RC, Gottesman II. Neurological soft signs as candidate endophenotypes for schizophrenia: A shooting star or a northern star? *Neurosci Biobehav Rev*. 2008;32(5):957-971.

51. Costain G, Bassett AS. Clinical applications of schizophrenia genetics: Genetic diagnosis, risk, and counseling in the molecular era. *Appl Clin Genet*. 2012;5:1-18.